Numerous similar tests
– a miracle cure in experimental life sciences?

Recently, an experimental physicist who had spent much of his long life doing basic medical research undertook to explain to me what is evidently a fundamental difference between a physicist's approach to experimentation and that of a researcher trained in the biosciences. "We who work in the exact natural sciences," he said, "plan and prepare our measurements with enormous precision, so as to keep track of as many as possible of the parameters that might affect the results of an experiment, and control them as accurately as modern technology allows. There is a reason for this, of course: if our efforts have met with success, replications of identical experiments will give results with negligible scatter, as far as the particular question at issue is concerned. Having established this base, when dealing with new but similar questions we no longer need to replicate the measurements several times, as long as the experimental situation is identical. To cite a trivial example: no one nowadays would think of recording NMR or IR spectra of a given substance many times in succession, unless there were some special considerations."

"You biomedical scientists," the elderly gentleman continued, "don't take your efforts to control the experimental situation seriously enough for my liking: very often, you do not check up all the parameters that can affect the outcome of an experiment as accurately as possible, and as a result you are often not as successful as you might otherwise be. Of course, I realize that the objects you study are mostly far more complex than is the case in many physical experiments, and so it seems to you impossible in principle to monitor all such parameters with sufficient accuracy. Investigation of the electrical activity of a neuron in a living, anaesthetized animal is simply not comparable in this respect with the apparently straightforward measurement of the electrical resistivity of copper, for example. That is why you usually have to repeat your measurements regarding a given question often enough that you can apply statistical methods, so that even though a broad scatter of the data is often unavoidable, the numerical results will be reliable to some degree. But keep in mind that you usually don't know the distribution of your data samples, and make sure not to be lulled into a false sense of security by your mean values, standard deviations and the related test procedures — for instance, when you have skewed distributions!"

While I was still thinking over what this advice might be likely to mean in practice for my current experimental approach, my companion evidently decided I needed some reassurance. He said, "The range of your experimental data is quite large, something like 60% of the median. So perhaps you weren't very successful in your attempts to control all the relevant parameters exactly. But don't worry about that yet; often a more accurate assessment of the range of a set of data cannot be obtained until exhaustively analyzing all the pertinent background information that is available about the measurement techniques used in your particular experiment. That is the reason why there are no generally valid rules for setting an upper limit on a potentially 'permissible' scatter of data!"

After a brief pause, he resumed. "But maybe your data suffer from something quite different. Apart from the more or less stochastic errors in measurement we have just been considering, can it be that you are also faced with systematic measurement errors? I'm just asking for the sake of completeness, because as you know, a standard deviation would tell us nothing on this score. In microelectrode experiments, for example, you would not know about such systematic errors until you really understood in detail all the interactions between the object you are studying and the recording electronics you are using — and even then you might not know everything! As P. B. Brown and his colleages justly complained not long ago (Brown et al. 1982), 'in the field of physiology, some reseachers talk of physiological control systems with positive or negative feedback, and are often unaware of the simplest principles of the control system theory' ".

"You see, my young friend, the value of repeating one and the same experiment many times is definitely limited. Much better to do everything to make the scatter of data as small as you possibly can, and this applies especially to the systematic disturbance variables, as difficult as these often are to discern. Spare no effort here! Only after both have been minimized to the full extent allowed by the currently available technology will you be able to expect results as reliable as is consistent with the present state of our knowledge. And those are the only results worth using to develop new hypotheses. As T. S. Kuhn (1961) put it: 'at best the criterion [for agreement between hypothesis and measurement] must be in agreement within the limits of accuracy of the measuring instruments employed' ". It was not until the other day, when my visitor had already departed, that the practical consequences of his advice occurred to me: quite possibly there would be considerable savings in experimental animals, time and hence in financial expense! How on earth could the old gentleman have forgotten that?

As you may have inferred from the above, the editors of GENERAL PHYSIOLOGY AND BIOPHYSICS do not regard large sample size as a mark of quality *per se*, and certainly not as an end in itself, but rather as a last resort for obtaining data that are worth reporting despite being scattered over a considerable range. In the sometimes extremely problematic area of systematic errors in measurement, however, they must rely on the special knowledge of their Field Editors and, in particular, of the referees they call upon for advice.

E. Koppenhöfer, Kiel